

(19)日本国特許庁 (J19)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平 1 1 - 1 7 5 5 1 7

(43)公開日 平成 1 1 年 ( 1 9 9 9 ) 7 月 2 日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G06F 17/22			G06F 15/20	503
3/16	320		3/16	320 H
17/60			H04N 7/15	
H04N 7/15			G06F 15/21	Z
// G06T 13/00			15/62	340 A

審査請求 未請求 請求項の数 8 O L (全 1 7 頁)

(21)出願番号 特願平 9 - 3 4 0 2 6 1

(22)出願日 平成 9 年 ( 1 9 9 7 ) 1 2 月 1 0 日

(71)出願人 0 0 0 0 0 1 0 0 7

キヤノン株式会社

東京都大田区下丸子 3 丁目 3 0 番 2 号

(72)発明者 齋藤 和之

東京都大田区下丸子 3 丁目 3 0 番 2 号 キ

ヤノン株式会社内

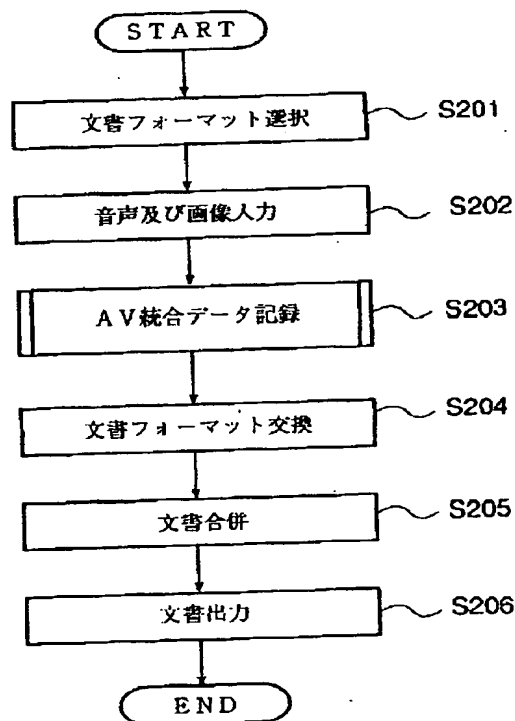
(74)代理人 弁理士 大塚 康徳 (外 2 名)

(54)【発明の名称】 情報処理装置およびその方法

(57)【要約】

【課題】 テレビ会議における議事録を生成する場合に、各情報の発生場所および発生順を示すことができる情報処理装置およびその方法を提供することを目的とする。

【解決手段】 ステップ S 2 0 1 では、文書フォーマットを選択する。ステップ S 2 0 2 では、サウンドデータおよび画像データを入力する。ステップ S 2 0 3 では、サウンドデータおよび画像データを所定の時間間隔毎にブロック化する。ステップ S 2 0 4 では、サウンドデータに基づいて作成されたテキストデータ、および、画像データを 1 ブロック毎に R T F 形式のフォーマットに変換する。ステップ S 2 0 5 では、R T F 形式に変換したテキストデータおよび画像データを、記録紙などに印刷する場合のレイアウトを設定する。ステップ S 2 0 6 では、ステップ S 2 0 5 で合成された文書データに対応する画像をプリンタ等から印刷する。



REST AVAILABLE COPY

## 【特許請求の範囲】

【請求項 1】 異なる場所から送られてくるそれぞれ異なる情報を入力する入力手段と、

前記入力情報に含まれる音声を認識してテキストを生成する生成手段と、

生成されるテキストと前記入力情報に含まれる画像とを、それらの情報の発生場所および発生順に基づき合成する合成手段とを有することを特徴とする情報処理装置。

【請求項 2】 前記テキストおよび前記画像は、合成された情報に基づき記録媒体上に可視像が形成される場合、情報の発生順に、発生場所の異なる情報が列方向に配置されるように合成されることを特徴とする請求項 1 に記載された情報処理装置。

【請求項 3】 前記生成手段は、前記入力情報に含まれる音声データを有音ブロックおよび無音ブロックに分割し、前記有音ブロックに含まれる音声を認識することを特徴とする請求項 1 または請求項 2 に記載された情報処理装置。

【請求項 4】 前記合成手段による合成処理は、前記有音ブロックに対応する動画像データから得られる静止画像データが表す画像に基づき行われることを特徴とする請求項 3 に記載された情報処理装置。

【請求項 5】 さらに、前記情報処理装置は、前記テキストのファイル形式を所定のファイル形式に変換する変換手段を備えることを特徴とする請求項 1 に記載された情報処理装置。

【請求項 6】 前記所定のファイル形式は、少なくとも HTML (HyperText Markup Language) 形式、RTF (Rich Text Format) 形式、PDL (Page Description Language) 形式の何れか 1 つであることを特徴とする請求項 5 に記載された情報処理装置。

【請求項 7】 異なる場所から送られてくるそれぞれ異なる情報を入力し、前記入力情報に含まれる音声を認識してテキストを生成し、

生成されるテキストと前記入力情報に含まれる画像とを、それらの情報の発生場所および発生順に基づき合成することを特徴とする情報処理方法。

【請求項 8】 情報処理のプログラムコードが記憶された記憶媒体であって、

異なる場所から送られてくるそれぞれ異なる情報を入力するステップのプログラムコードと、

前記入力情報に含まれる音声を認識してテキストを生成するステップのプログラムコードと、

生成されるテキストと前記入力情報に含まれる画像とを、それらの情報の発生場所および発生順に基づき合成するステップのプログラムコードとを有することを特徴とする記憶媒体。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、情報処理装置およびその方法に関し、例えば、入力情報に基づいて、会議の議事録などを作成することができる情報処理装置およびその方法に関するものである。

【0002】

【従来の技術】 近年、ネットワークを利用して距離的に離れた各会議室を結んで会議を行うという、いわゆるテレビ会議が一般化しつつある。このテレビ会議では、距離的に離れた各会議室にいる相手の容姿や発言は、会議室に設置されたビデオカメラなどを通して配信される。また、音声認識および画像認識技術を応用すれば、記録された画像および音声に基づいて、会議の内容が記録された議事録を作成することも考えられる。

【0003】

【発明が解決しようとする課題】 しかし、上述した技術においては、次のような問題がある。

【0004】 音声認識および画像認識を用いて作成される議事録は、ある程度会議の流れにそった形で発言内容をまとめることができる。しかし、その議事録からそれぞれの発言がどの会議室で発言されたものかを判断することは容易にできない。

【0005】 本発明は、上記の問題を解決するものであり、例えば、テレビ会議における議事録を生成する場合に、各情報の発生場所および発生順を示すことができる情報処理装置およびその方法を提供することを目的とする。

【0006】

【課題を解決するための手段】 本発明は、前記の目的を達成する一手段として、以下の構成を備える。

【0007】 本発明にかかる情報処理装置は、異なる場所から送られてくるそれぞれ異なる情報を入力する入力手段と、前記入力情報に含まれる音声を認識してテキストを生成する生成手段と、生成されるテキストと前記入力情報に含まれる画像とを、それらの情報の発生場所および発生順に基づき合成する合成手段とを有することを特徴とする。

【0008】 本発明にかかる情報処理方法は、異なる場所から送られてくるそれぞれ異なる情報を入力し、前記入力情報に含まれる音声を認識してテキストを生成し、生成されるテキストと前記入力情報に含まれる画像とを、それらの情報の発生場所および発生順に基づき合成することを特徴とする。

【0009】

【発明の実施の形態】 以下、本発明にかかる一実施形態の情報処理装置およびその方法について図を参照して詳細に説明する。

【0010】 【第一実施形態】 図 1 は、本発明にかかる一実施形態の情報処理装置の構成例を示すブロック図で

ある。

【0011】本発明の情報処理装置は画像入力部101、画像入力インタフェイス部102、音声入力部103、音声入力インタフェイス部104、CPU105、ROM106、RAM107、出力部108、出力インタフェイス部109、外部記憶部110、外部記憶インタフェイス部111、システムバス112を備える。

【0012】画像入力部101は、外部機器などから画像を入力するための画像入力部である。画像入力インタフェイス部102は、画像入力部101とシステムバス112とをインタフェイスする。音声入力部103は、外部機器などから音声等の音声を入力するための音声入力部である。音声入力インタフェイス部104は、音声入力部103とシステムバス112とをインタフェイスする。CPU105は、装置全体の制御および本発明のアプリケーションプログラム等を実行するためのCPUである。

【0013】ROM106は、装置全体を制御するためのプログラムや本発明のアプリケーションプログラム等を格納するためのROMである。RAM107は、入力された画像データや音声データの展開領域、CPU105の作業領域等として使用するためのRAMである。出力部108は、本発明の情報処理装置を使用して処理された文書データ等をモニタ等に出力するための出力部である。出力インタフェイス部109は出力部108とシステムバス112とをインタフェイスする。

【0014】外部記憶部110は、各種データ等を記憶するためのフロッピディスクやハードディスク等の外部記憶部である。外部記憶インタフェイス部111は、外部記憶部110とシステムバス112とをインタフェイスする。システムバス112は、画像入力インタフェイス部102、音声入力インタフェイス部104、CPU105、ROM106、RAM107、出力インタフェイス部109、外部記憶インタフェイス部111にインタフェイスする。

【0015】図2は、本発明にかかる第一実施形態の情報処理装置の処理について説明するためのフローチャートである。

【0016】なお、第一実施形態では、例えば、2地点間で行われるテレビ会議に本発明を適用した場合について説明を行う。

【0017】ステップS201では、入力された音声および動画像等を、HTML(Hyper Text Markup Language)形式、RTF(Rich Text Format)形式、PDL(Page Description Language)形式等の文書フォーマットに変換するために、ユーザが所望する文書フォーマット(ファイル形式)を選択する。第一実施形態では、例えば、RTF形式の文書フォーマットがユーザにより選択されたものとして、以下の説明を行う。

【0018】ステップS202では、例えば、ネットワーク等に接続された外部機器等から出力される音声および動画像等を、音声入力部103および画像入力部101を介して入力する。第一実施形態では、2地点間で行われるテレビ会議を例としているので、このステップS202では、例えば、地点A側に備付けられた外部機器から出力される音声および動画像等と、地点B側に備付けられた外部機器から出力される音声および動画像等とが入力される。

【0019】ステップS203では、入力された音声データに基づき、音声データを有音ブロックおよび無音ブロックに分割し、オーディオ・ビジュアル統合データ(以後「AV統合データ」とする)として、AV統合データ格納部41に記録する。また、入力された音声や動画像データの発生場所を示す位置情報や時間情報などを外部機器などから入力し、AV統合データ格納部41に含まれる位置情報および時間情報を記憶するためのエリアに記録する。このステップS203の詳細な処理については、図3のフローチャートに従って説明する。

【0020】ステップS301では、入力された音声データに基づき、音声データを有音ブロックおよび無音ブロックに分割する。上述したブロック化について、図5を用いて説明すると次のようになる。つまり、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間以上の場合、その期間に対応する音声データを1ブロックの有音ブロックとし、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間未満の期間、または、音声データの再生レベルが所定のレベル未満になってから所定のレベルに達するまでの期間に対応する音声データを1ブロックの無音ブロックとする。

【0021】ステップS302では、一ブロック単位に音声データの有音ブロックに対する音声認識を行ない、音声データからテキストコードへの変換を行なう。なお、この音声認識処理は周知の技術により実行可能である。ステップS303では、図4に示すように、コード変換した一ブロック分のテキストデータをAV統合データ格納部41に含まれるテキストデータを記憶するためのエリアに記録する。

【0022】ステップS304では、音声データの一ブロック分の有音ブロックに対応する動画像データを所定の時間間隔T毎で区切り、区切った時間に対応する動画像データの一部をn個の静止画像データとして選択する。そして、そのn個の静止画像データを有効画像データとし、音声データの1ブロック分の無音ブロックに対応する動画像データを無効画像データとする。

【0023】上述したように動画像データの一部をn個の静止画像データとして選択する場合、例えば、システムあるいはユーザなどにより、動画像データの最初のフ

フレームに対応する静止画像データのみを選択するようにしてもよい。このようにすれば、記憶域に余分なデータが記憶されないため、記憶域を有効に使用できるなどの効果を得ることができる。

【0024】ステップS305では、図4に示すように、一ブロック分のn個の有効画像データを、AV統合データ格納部41に含まれる静止画像（有効画像）データを記憶するためのエリアに記録する。

【0025】ステップS306では、AV統合データ格納部41に格納された一ブロック分の静止画像データに含まれる文字や表などを画像認識し、画像認識データに変換する。なお、この画像認識処理は周知の技術により実行可能である。ステップS307では、図4に示すように、変換した一ブロック分の画像認識データを、AV統合データ格納部41に含まれる画像認識データを記憶するためのエリアに記録する。

【0026】ステップS308では、全ブロックに対して、ステップS302～ステップS307までの処理が終了したか否か判断し、終了した場合ステップS204へ進み、終了していない場合ステップS302へ戻る。

【0027】以上のようにして、各々のブロックに対応する音声データおよび動画データに基づいて生成されたテキストデータ、静止画像（有効画像）データ、画像認識データは、AV統合データ格納部41に含まれるそれぞれの記憶エリアに記録される。

【0028】ステップS204では、AV統合データ格納部41に格納されたテキストデータを1ブロック毎にRTF形式のフォーマットに変換し、さらに、テキストデータに対応するn個の静止画像データをRTF形式に適した画像形式に変換する。このステップでは、RTF形式に適した画像形式に変換する静止画像データをn個の静止画像データの中から任意に選択することもできる。

【0029】ステップS205では、ユーザあるいは装置により、例えば、RTF形式に変換したテキストデータおよび静止画像データ（あるいは、画像認識データ）を、プリンタ等を使用して記録媒体に記録する際のレイアウトが設定される。つまり、第一実施形態では、2地点間で行われるテレビ会議を例としているので、このステップS205では、例えば、会議中に地点A側で発言された内容等を示す音声データおよび地点A側の会議室の様子等を示す動画データに基づいて、RTF形式に変換したテキストデータおよび静止画像データ（あるいは、画像認識データ）と、会議中に地点B側で発言された内容等を示す音声データおよび地点B側の会議室の様子等を示す動画データに基づいて、RTF形式に変換したテキストデータ、および、静止画像データ（あるいは、画像認識データ）とを合成し文書データ化するという処理を行う。なお、第一実施形態では、複数の静止画像データ全てを文書データ化しているが、本発明はこれ

に限られたものではなく、例えば、複数の静止画像データの中から任意に静止画像データを選択して、文書データ化することも可能である。

【0030】ステップS206では、ステップS205で合成された文書データに対応する画像をプリンタ等から印刷する。

【0031】図6は、合成された文書データの印刷例を示す図である。

【0032】7001は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。7002は、地点A側の会議室の様子等を示す動画データに基づいて生成された静止画像データ、あるいは、画像認識データである。7003は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。7004は、地点B側の会議室の様子等を示す動画データに基づいて生成された静止画像データ、あるいは、画像認識データである。

【0033】7005に示すように、地点A側の音声データおよび動画データに基づいて生成されたテキストデータおよび静止画像データと、地点B側の音声データおよび動画データに基づいて生成されたテキストデータおよび静止画像データとを、時系列（直列）に配置すると会議の流れにそった形で会議の議事録などを作成することができる。

【0034】また、図7の6001は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。6002は、地点A側の会議室の様子等を示す動画データに基づいて生成された静止画像データである。6003は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。6004は、地点B側の会議室の様子等を示す動画データに基づいて生成された静止画像データである。

【0035】6005に示すように、地点A側の音声データおよび動画データに基づいて生成されたテキストデータおよび静止画像データを例えば記録紙等の左側に配置し、また、地点B側の音声データおよび動画データに基づいて生成されたテキストデータおよび静止画像データを例えば記録紙等の右側に配置する。さらに、発言の順番に縦方向にテキストデータおよび画像データを配置すれば、会議の流れにそった形で会議の議事録などを作成することができる。

【0036】〔第二実施形態〕第二実施形態では、AV統合データ格納部41に格納された静止画像データに含まれる文字や表などを画像認識し、会議中に使用された資料等をテキストコードや表データへ変換して、会議の議事録などを作成することができる。

【0037】この第二実施形態の情報処理装置の処理について、図2および図8のフローチャートに従って説明する。

【0038】図8は、図2のステップS203の詳細な処理について説明するためのフローチャートである。なお、第二実施形態では、図2のステップS203の処理のみが第一実施形態の処理と異なるため、図2のステップS201、ステップS202、ステップS204～ステップS206については詳細な説明を省略する。

【0039】ステップS801では、入力された音声データに基づき、音声データを有音ブロックおよび無音ブロックに分割する。上述したブロック化について、図5を用いて説明すると次のようになる。つまり、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間以上の場合、その期間に対応する音声データを1ブロックの有音ブロックとし、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間未満の期間、または、音声データの再生レベルが所定のレベル未満になってから所定のレベルに達するまでの期間に対応する音声データを1ブロックの無音ブロックとする。

【0040】ステップS802では、一ブロック単位に音声データの音声認識を行ない、音声データからテキストコードへの変換を行なう。なお、この音声認識処理は周知の技術により実行可能である。ステップS803では、図9に示すように、コード変換した一ブロック分のテキストデータを、AV統合データ格納部41に含まれるテキストデータを記憶するためのエリアに記録する。

【0041】ステップS804では、音声データの一ブロック分の有音ブロックに対応する動画像データを所定の時間間隔T毎に分割してn個の静止画像データを選択する。そして、そのn個の静止画像データを有効画像データとし、音声データの1ブロック分の無音ブロックに対応する動画像データを無効画像データとする。ステップS805では、図9に示すように、一ブロック分のn個の静止画像（有効画像）データを、AV統合データ格納部41に含まれる静止画像（有効画像）データを記憶するためのエリアに記録する。

【0042】ステップS806では、AV統合データ格納部41に格納された一ブロック分の静止画像データに含まれる文字や表などを画像認識し、静止画像データに文字領域が含まれている場合、文字認識処理によりテキストコードに変換する。なお、この文字認識処理は周知の技術により実行可能である。ステップS807では、図9に示すように、変換した一ブロック分のテキストデータを、AV統合データ格納部41に含まれる文字認識データを記憶するためのエリアに記録する。

【0043】ステップS808では、AV統合データ格納部41に格納された一ブロック分の静止画像データに含まれる文字や表などを画像認識し、静止画像データに表領域が含まれている場合、例えば、表計算ソフトウェアなどで使用することができる表データ形式に変換す

る。なお、この画像認識処理は周知の技術により実行可能である。ステップS809では、図9に示すように、変換した一ブロック分の表データを、AV統合データ格納部41に含まれる表データを記憶するためのエリアに記録する。

【0044】ステップS810では、一ブロック分の文字認識データn個の静止画像データのうち、ほぼ同じ内容の画像データが存在する場合は1つを残し他を無効と判断し廃棄する。例えば、2つの静止画像データの差分をとり、その差分が所定の範囲内である場合、2つの静止画像データはほぼ同じ静止画像データであると判断することができる。

【0045】ステップS811では、AV統合データ格納部41にテキストデータが格納されている場合、図9のAV統合データ格納部41に含まれる有効フラグの内容を記録するためのエリアに、例えば「1」を記録する。また、AV統合データ格納部41にテキストデータが格納されていない場合、図9のAV統合データ格納部41に含まれる有効フラグの内容を記録するためのエリアに、例えば「0」を記録する。ステップS812では、全ブロックに対して、ステップS802～ステップS811までの処理が終了したか否か判断し、終了した場合ステップS813へ進み、終了していない場合ステップS802へ戻る。

【0046】ステップS813では、全ブロックのAV統合データ格納部41に含まれる有効フラグを参照し、有効フラグの内容が「1」であるブロックを有効ブロックとして選択する。そして、ステップS204へ進み、選択された有効ブロックに含まれるテキストデータを1ブロック毎にRTF形式のフォーマットに変換し、さらに、テキストデータに対応するn個の静止画像データをRTF形式に適した画像形式に変換する。

【0047】以上のようにして、各々のブロックに対応する音声データおよび動画像データに基づいて生成されたテキストデータ、静止画像（有効画像）データ、表データ、文字認識データ、AV統合データ格納部41に含まれるそれぞれの記憶エリアに記録される。

【0048】図10は、合成された文書データの印刷例を示す図である。

【0049】10001は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。10002は、地点A側の会議室の様子等を示す動画像データに基づいて生成された静止画像データである。10003は、地点A側で発言された内容等を示す音声データに基づいて生成されたテキストデータである。10004は、地点B側の会議室の様子等を示す動画像データに基づいて生成された静止画像データである。10006は、静止画像データ10002を画像認識することにより生成された画像認識データである。

【0050】10005に示すように、地点A側の音声

10

20

30

40

50

データおよび動画像データに基づいて生成されたテキストデータおよび静止画像データを例えば記録紙等の左側に配置し、また、地点 B 側の音声データおよび動画像データに基づいて生成されたテキストデータおよび静止画像データを例えば記録紙等の右側に配置し、そして、画像認識データ 1 0 0 0 6 を記録紙の任意の位置に配置すると、会議の流れにそった形でさらに詳しい会議の議事録などを作成することができる。

【0051】【第三実施形態】第三実施形態では、HTML 形式の文書フォーマットにより動画像データを取扱う場合について説明する。

【0052】図 1 1 は、本発明にかかる第三実施形態の情報処理装置の処理について説明するためのフローチャートである。

【0053】ステップ S 1 0 1 では、例えば、ネットワーク等に接続された外部機器等から出力される音声および動画像を、音声入力部 1 0 3 および画像入力部 1 0 1 を介して入力する。ステップ S 1 0 2 では、入力された音声データに基づき、音声データを有音ブロックおよび無音ブロックに分割する。上述したブロック化について、図 1 3 を用いて説明すると次のようになる。つまり、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間以上の場合、その期間に対応する音声データを 1 ブロックの有音ブロックとし、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間未満の期間、または音声データの再生レベルが所定のレベル未満になってから所定のレベルに達するまでの期間に対応する音声データを 1 ブロックの無音ブロックとする。

【0054】ステップ S 1 0 3 では、一ブロック単位に音声データの有音ブロックに対する音声認識を行ない、音声データからテキストコードへの変換を行なう。なお、この音声認識処理は周知の技術により実行可能である。ステップ S 1 0 4 では、図 1 2 に示すようにコード変換した一ブロック分のテキストデータを、AV 統合データ格納部 1 2 1 に含まれるテキストデータを記憶するためのエリアに記録する。

【0055】ステップ S 1 0 5 では、音声データの一ブロック分の有音ブロックに対応する動画像データを有効画像データとし、音声データの 1 ブロック分の無音ブロックに対応する動画像データを無効画像データとする。ステップ S 1 0 6 では、図 1 2 に示すように一ブロック分の動画像（有効画像）データを、AV 統合データ格納部 1 2 1 に含まれる動画像（有効画像）データを記憶するためのエリアに記録する。

【0056】ステップ S 1 0 7 では、全ブロックに対して、ステップ S 1 0 3 ～ステップ S 1 0 6 までの処理が終了したか否かを判断し、終了した場合ステップ S 1 0 8

へ進み、終了していない場合ステップ S 1 0 3 へ戻る。ステップ S 1 0 8 では、入力された音声および動画像等を、HTML 形式、RTF 形式、PDL 形式等の文書フォーマットに変換するために、ユーザが所望する文書フォーマットが選択される。第三実施形態では、例えば、HTML 形式の文書フォーマットがユーザにより選択されたものとする。

【0057】ステップ S 1 0 9 では、AV 統合データ格納部 1 2 1 に格納されたテキストデータを 1 ブロック毎に HTML 形式のフォーマットに変換し、さらに、テキストデータに対応する動画像データを HTML 形式に適した画像形式（例えば、AVI ファイルなど）に変換してファイルとして保存する。このように、HTML 形式の文書フォーマットを用いると、例えば、ブラウザなどを利用してモニタ上に表示された文字列をマウスなどでクリックすると、そのクリックした文字列に対応する動画像がモニタ上に表示されるなどの効果を得ることができる。

【0058】また、ステップ S 1 1 0 に示すように、ステップ S 1 0 9 で作成された HTML 形式の文書データに対応する画像をプリンタ等から印刷することもできる。

【0059】【第 4 実施形態】第四実施形態では、HTML 形式の文書フォーマットにより音声データを取扱う場合について説明する。

【0060】図 1 4 は、本発明にかかる第四実施形態の情報処理装置の処理について説明するためのフローチャートである。

【0061】ステップ S 7 0 1 では、例えば、ネットワーク等に接続された外部機器等から出力される音声および動画像を、音声入力部 1 0 3 および画像入力部 1 0 1 を介して入力する。

【0062】ステップ S 7 0 2 では、入力された音声データに基づき、音声データを有音ブロックおよび無音ブロックに分割する。上述したブロック化について、図 1 3 を用いて説明すると次のようになる。つまり、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間以上の場合、その期間に対応する音声データを 1 ブロックの有音ブロックとし、音声データの再生レベルが所定のレベルに達してから所定のレベル未満になるまでの時間があらかじめ設定された所定の時間未満の期間、または、音声データの再生レベルが所定のレベル未満になってから所定のレベルに達するまでの期間に対応する音声データを 1 ブロックの無音ブロックとする。

【0063】ステップ S 7 0 3 では、図 1 5 に示すように、音声データを、AV 統合データ格納部 1 5 1 に含まれる音声データを記憶するためのエリアに記録する。ステップ S 7 0 4 では、一ブロック単位に音声データの有音ブロックに対する音声認識を行ない、音声データから

テキストコードへの変換を行なう。なお、この音声認識処理は周知の技術により実行可能である。

【0064】ステップS705では、図15に示すように、コード変換した一ブロック分のテキストデータを、AV統合データ格納部151に含まれるテキストデータを記憶するためのエリアに記録する。ステップS706では、音声データの一ブロック分の有音ブロックに対応する動画像データを所定の時間間隔T毎に分割してn個の静止画像データを選択する。そして、そのn個の静止画像データを有効画像データとし、音声データの1ブロック分の無音ブロックに対応する動画像データを無効画像データとする。

【0065】ステップS707では、図15に示すように、一ブロック分のn個の静止画像（有効画像）データを、AV統合データ格納部151に含まれる静止画像（有効画像）データを記憶するためのエリアに記録する。ステップS708では、全ブロックに対して、ステップS703～ステップS707までの処理が終了したか否かを判断し、終了した場合ステップS709へ進み、終了していない場合ステップS703へ戻る。

【0066】ステップS709では、入力された音声および動画像等を、HTML形式、RTF形式、PDL形式等の文書フォーマットに変換するために、ユーザが所望する文書フォーマットが選択される。第四実施形態では、例えば、HTML形式の文書フォーマットがユーザにより選択されたものとする。

【0067】ステップS710では、AV統合データ格納部121に格納されたテキストデータを1ブロック毎にHTML形式のフォーマットに変換し、さらに、テキストデータに対応する音声データをHTML形式に適した音声形式（例えば、WAVファイルなど）に変換してファイルとして保存する。このように、HTML形式の文書フォーマットを用いると、例えば、ブラウザなどを利用してモニタ上に表示された文字列をマウスなどでクリックすると、そのクリックした文字列に対応する音声

がスピーカなどから出力されるという効果を得ることができる。

【0068】また、ステップS711に示すように、ステップS710で作成されたHTML形式の文書データに対応する画像をプリンタ等から印刷することもできる。

【0069】このように本発明によれば、例えば、ビデオカメラで撮影した映像や音声等に基づいて、音声認識処理、画像認識処理等を行うことにより撮影した映像や音声等を文書化することができる。また、同様にテレビなどから出力される音声や画像などを文書化することもできる。

【0070】

【他の実施形態】なお、本発明は、複数の機器（例えばホストコンピュータ、インタフェイス機器、リーダ、プ

リンクなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【0071】また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0072】この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【0073】プログラムコードを供給するための記憶媒体としては、例えば、フロッピーディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどを用いることができる。

【0074】また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0075】さらに、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0076】

【発明の効果】以上説明したように、本発明によれば、例えば、テレビ会議における議事録を生成する場合に、各情報の発生場所および発生順を示すことができる情報処理装置およびその方法を提供することができる。

【0077】

【図面の簡単な説明】

【図1】 本発明にかかる一実施形態の情報処理装置の構成例を示すブロック図、

【図2】 本発明にかかる一実施形態の情報処理装置の処理について説明するためのフローチャート、

【図3】 図2のステップS203の処理を詳細に説明するためのフローチャート、

【図4】 AV統合データ格納部の構造例を示す図、

【図5】 入力される音声および動画像のブロック化を説明するための図、

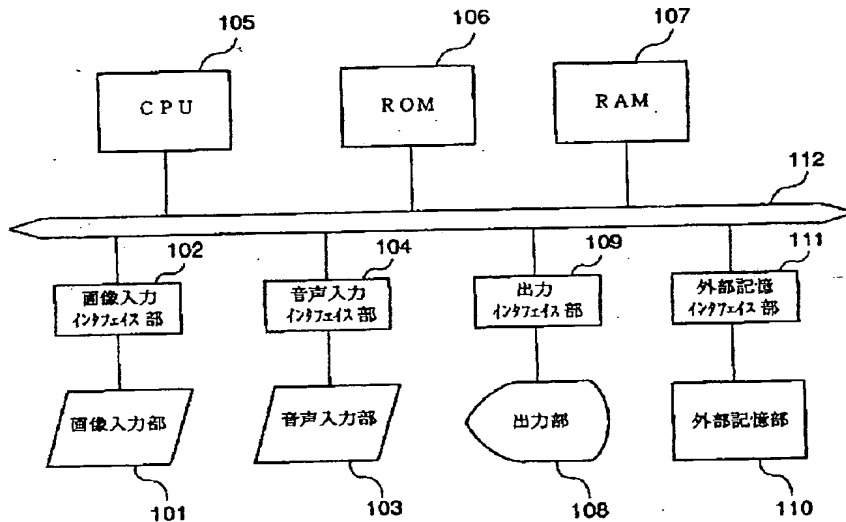
13

- 【図 6】 合成した文書の印刷例を示す図、  
 【図 7】 合成した文書の印刷例を示す図、  
 【図 8】 図 2 のステップ S 2 0 3 の処理を詳細に説明するためのフローチャート、  
 【図 9】 AV 統合データ格納部の構造例を示す図、  
 【図 1 0】 合成した文書の印刷例を示す図、  
 【図 1 1】 本発明にかかる一実施形態の情報処理装置の処理について説明するためのフローチャート、

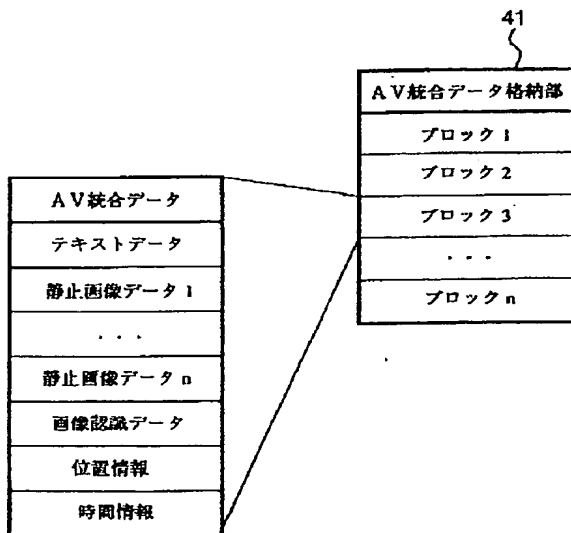
14

- 【図 1 2】 合成した文書の印刷例を示す図、  
 【図 1 3】 入力される音声および動画像のブロック化を説明するための図、  
 【図 1 4】 本発明にかかる一実施形態の情報処理装置の処理について説明するためのフローチャート、  
 【図 1 5】 AV 統合データ格納部の構造例を示す図である。

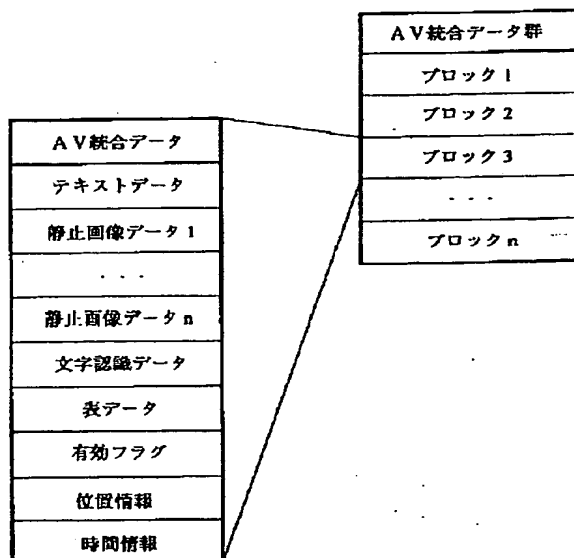
【図 1】



【図 4】

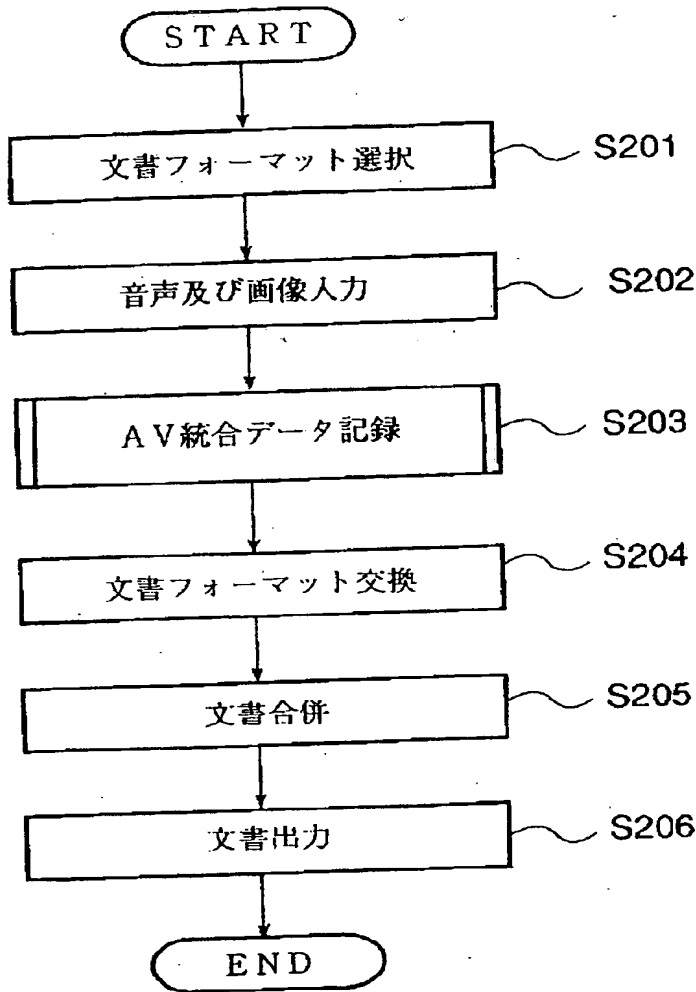


【図 9】

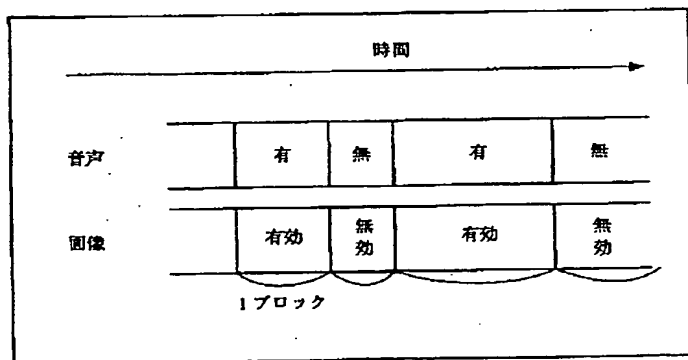




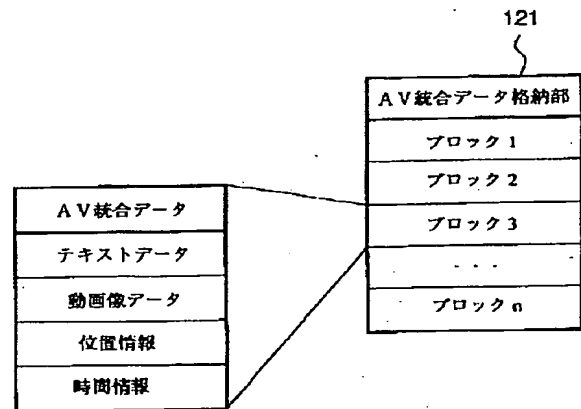
【図 2】



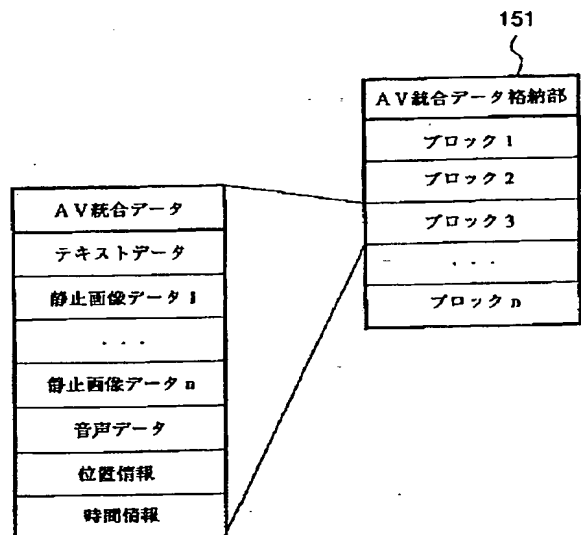
【図 13】



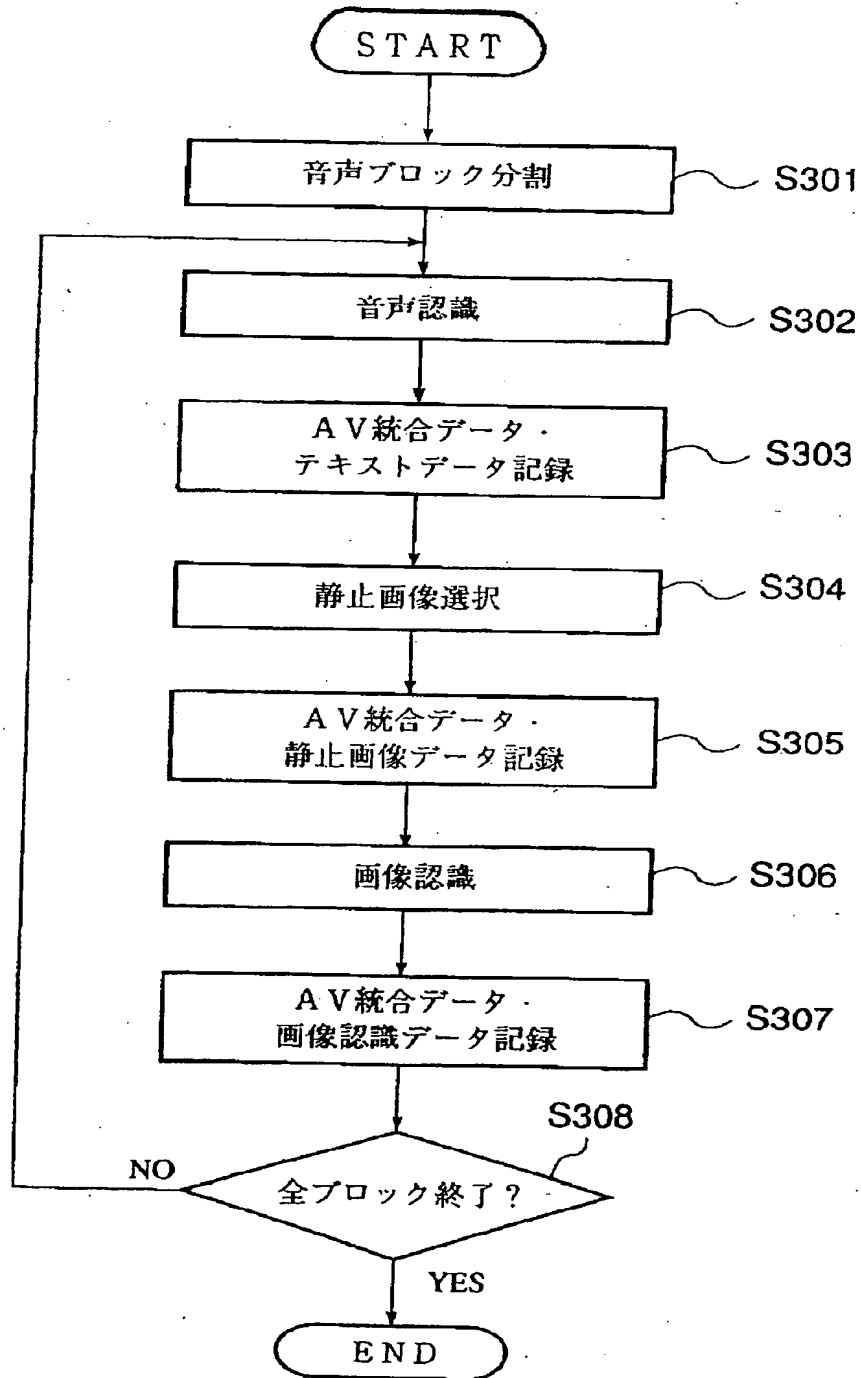
【図 12】



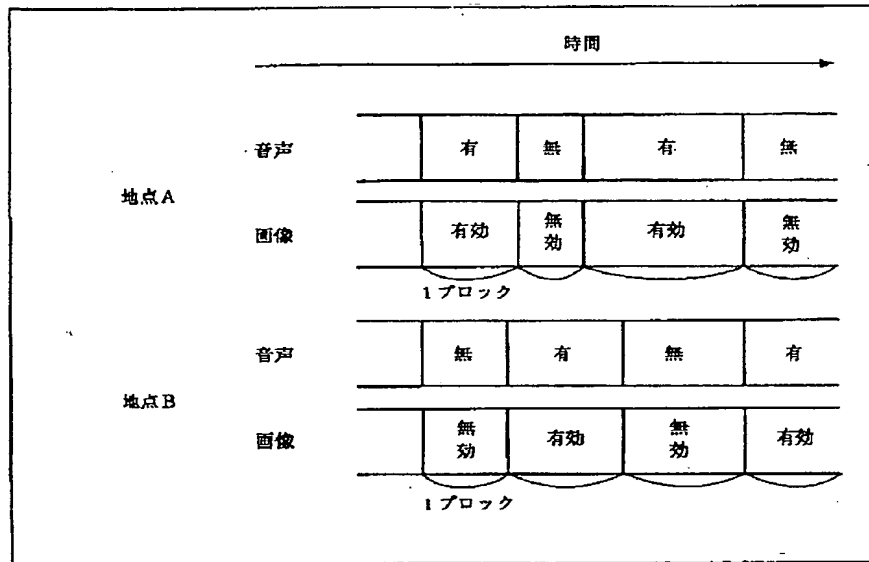
【図 15】



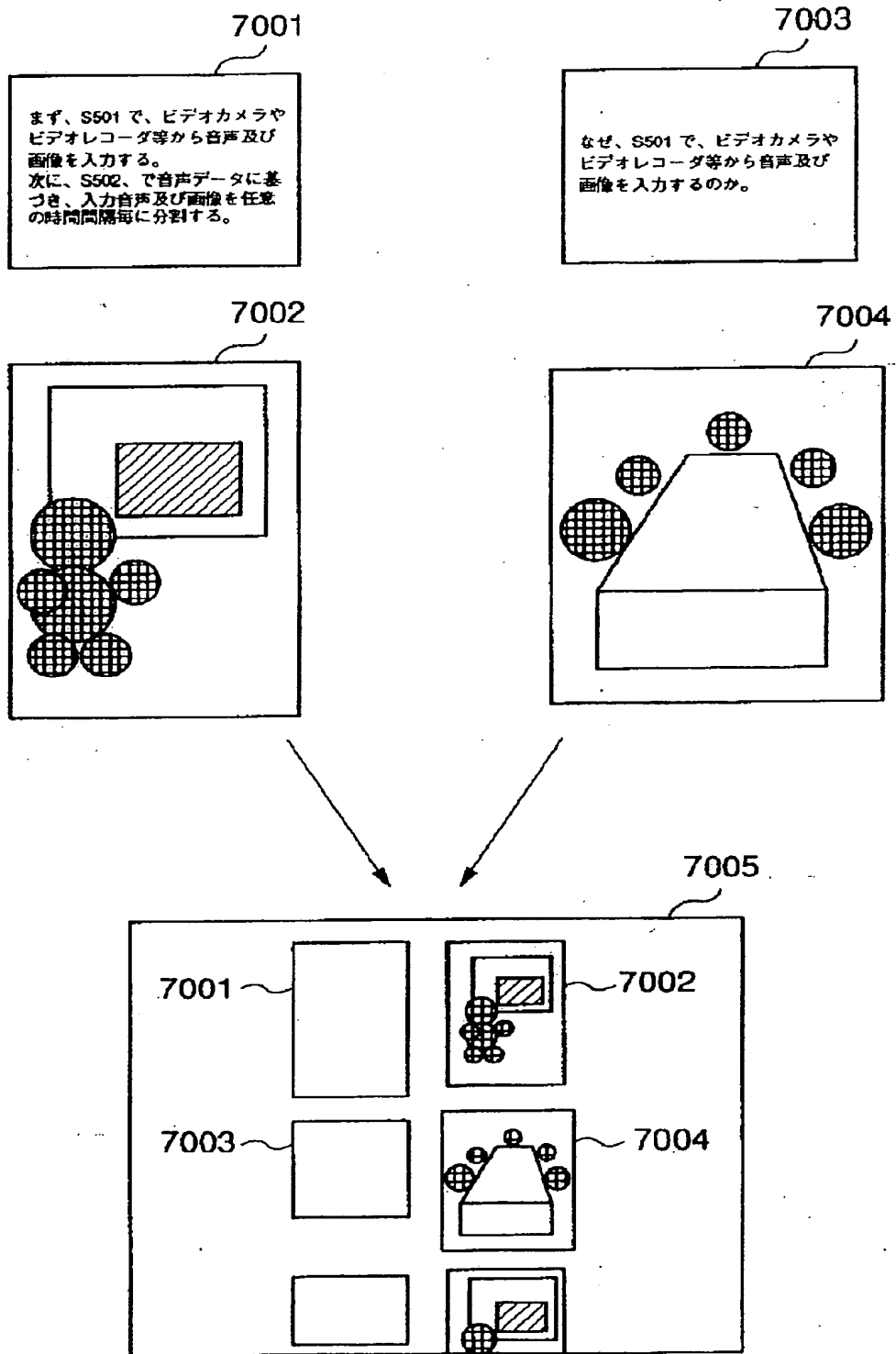
【図 3】



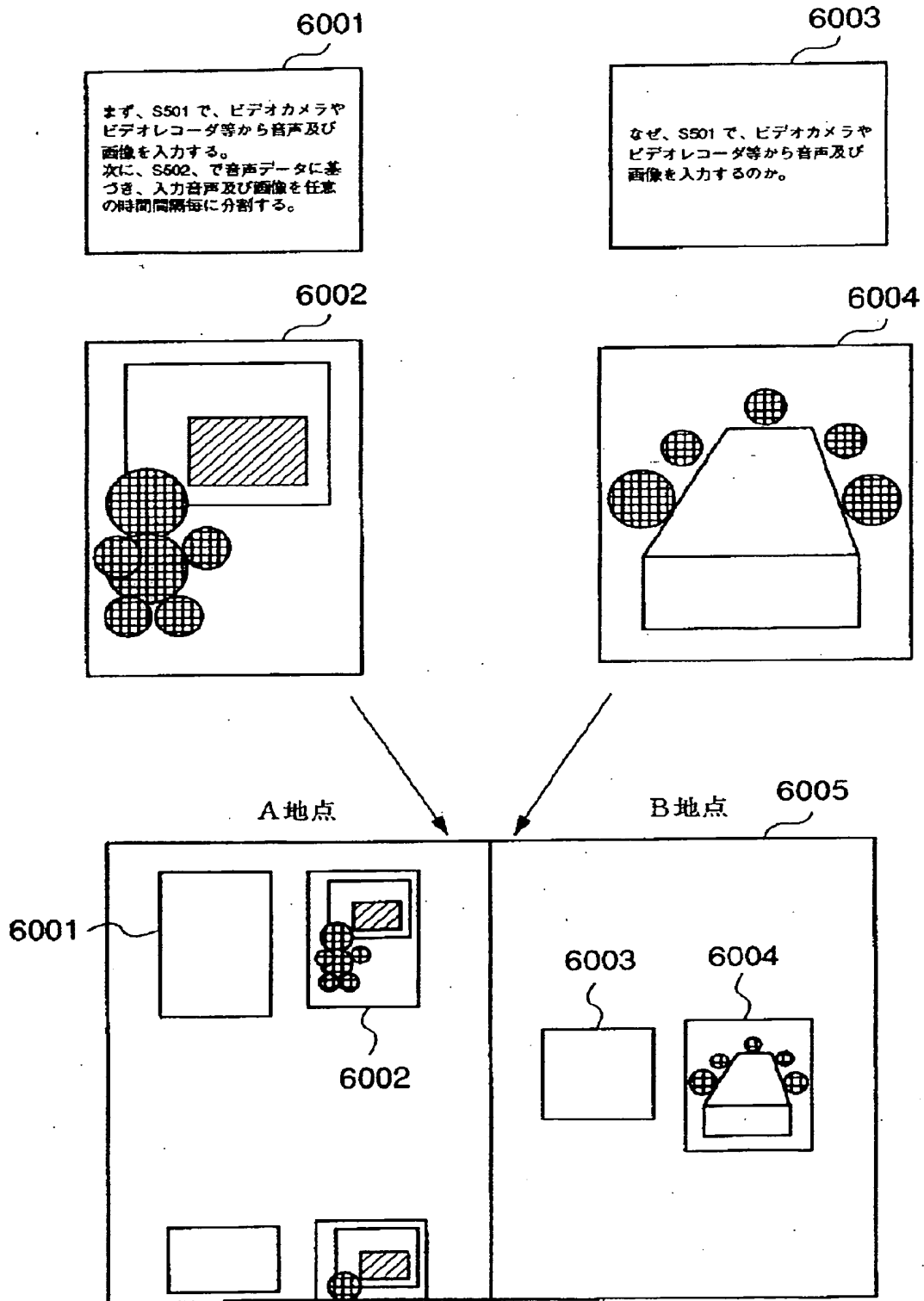
【図 5】



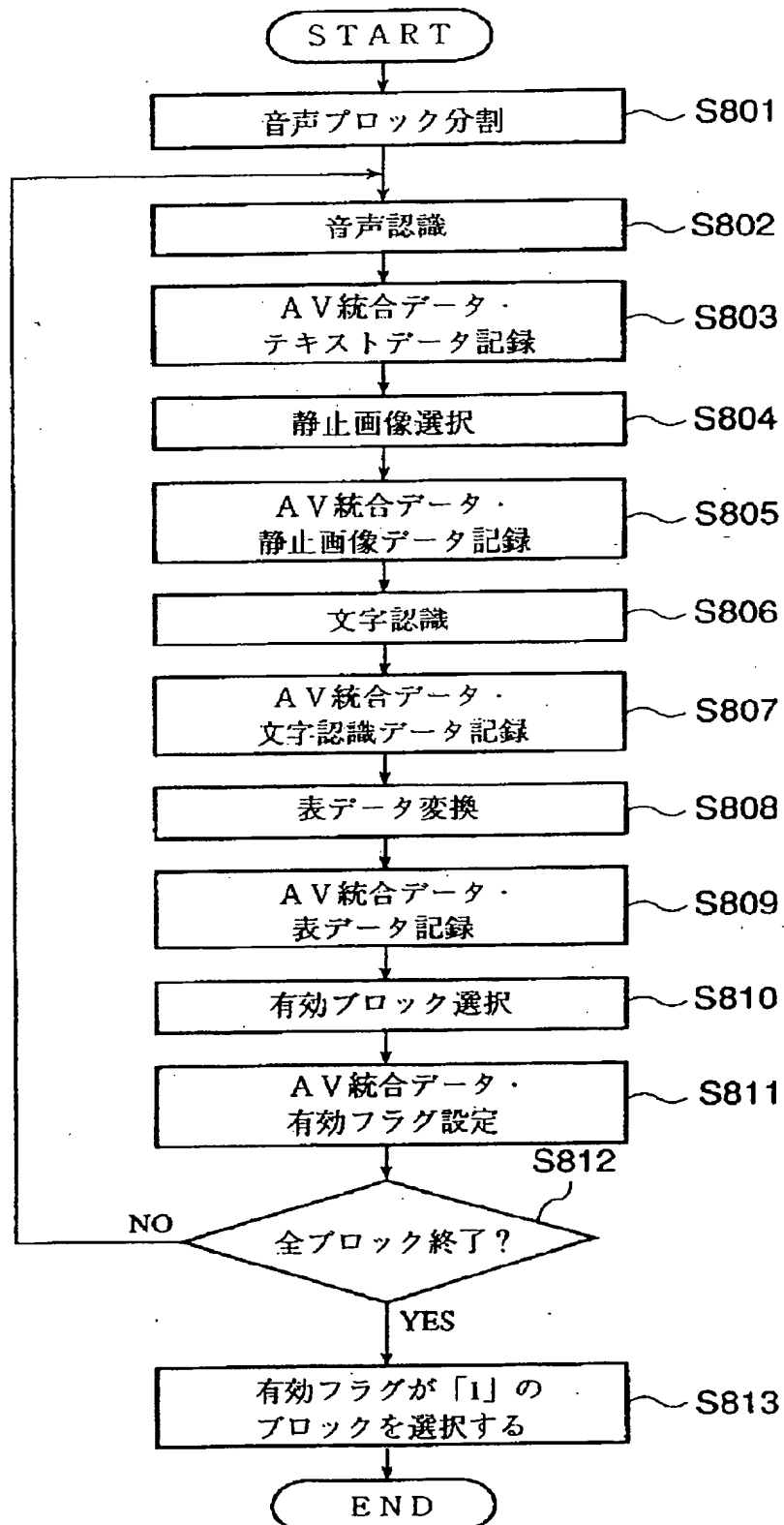
【図 6】



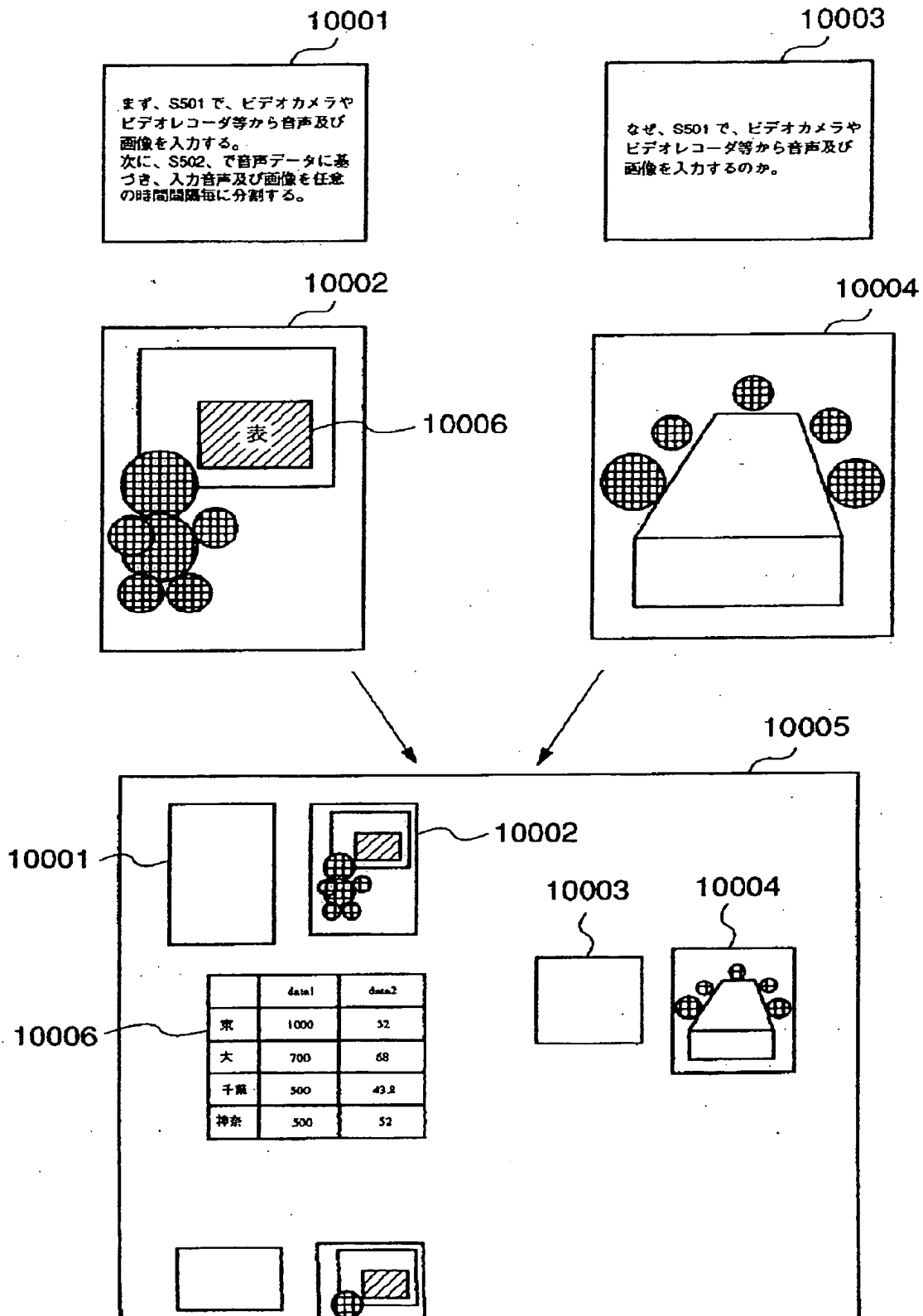
【 図 7 】



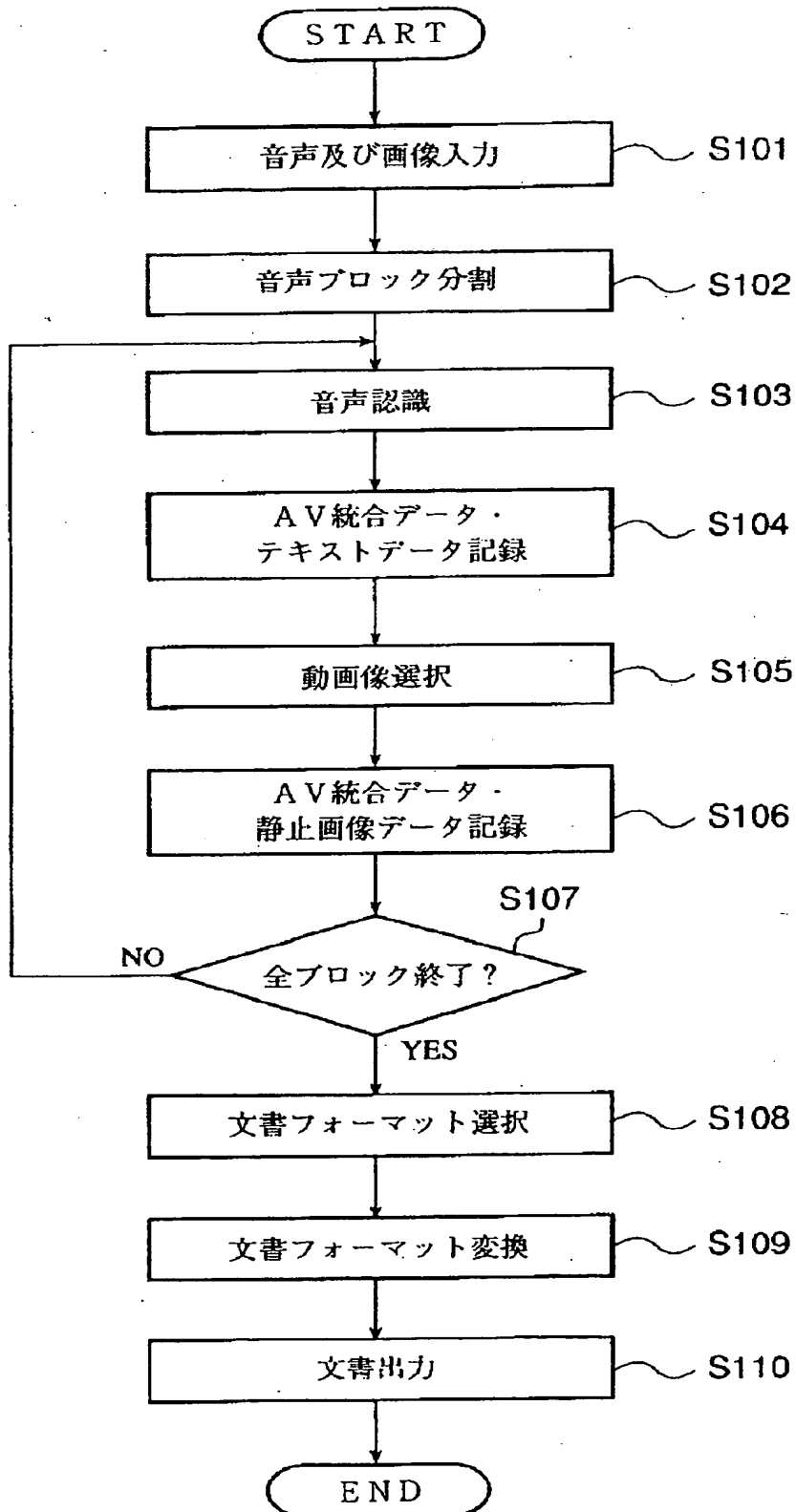
【図 8】



【図 10】

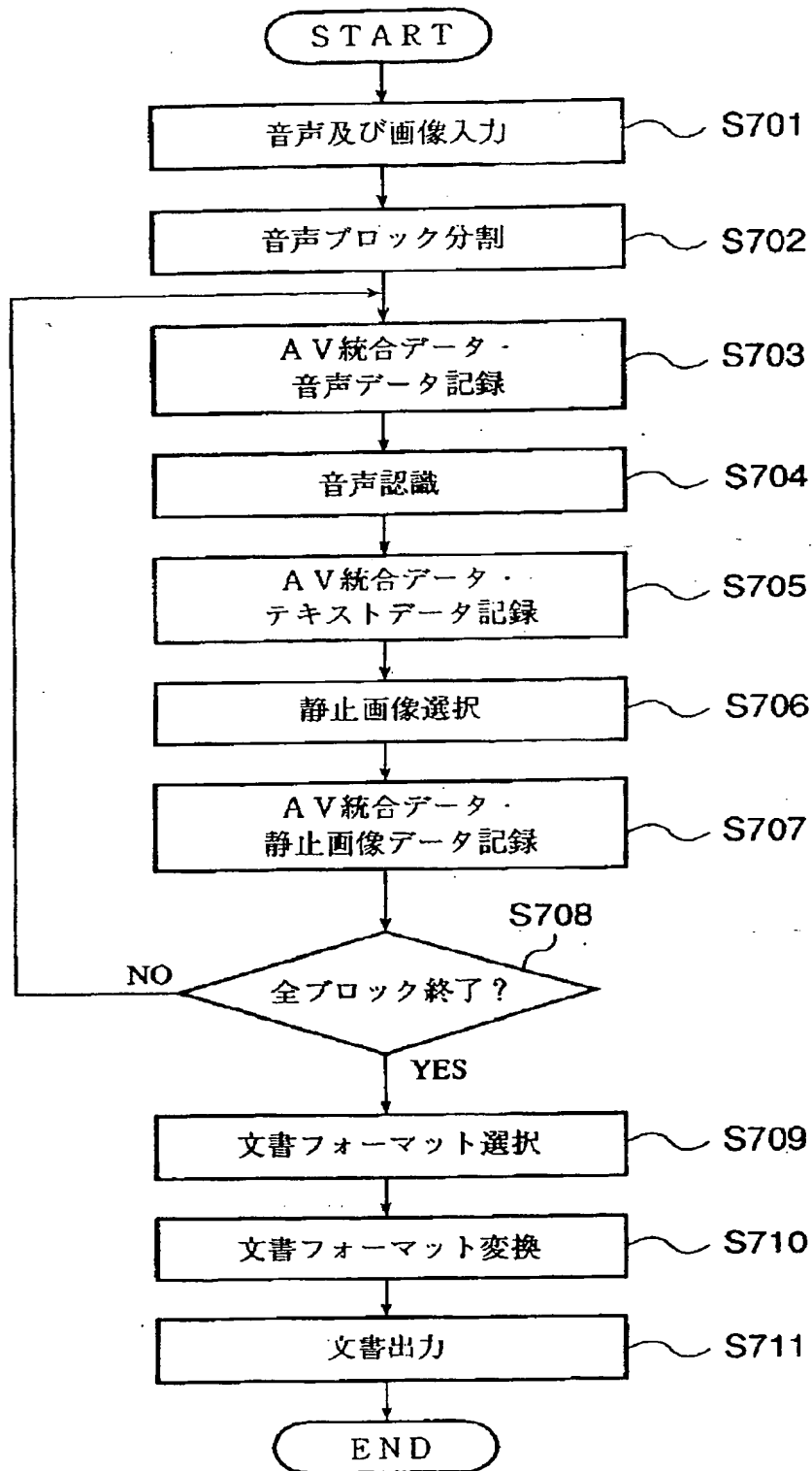


【図 11】





【図 14】



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**